


# Microsoft Academic Graph data schema

12/06/2018 4 minutes to read Contributors 

## In this article

[Open Data License: ODC-BY](#)

[Entity Relationship Diagram](#)

[Note on "rank"](#)

[Affiliations.txt](#)

[Authors.txt](#)

[ConferenceInstances.txt](#)

[ConferenceSeries.txt](#)

[FieldsOfStudyChildren.txt](#)

[FieldsOfStudy.txt](#)

[Journals.txt](#)

[PaperAbstractInvertedIndex.txt](#)

[PaperAuthorAffiliations.txt](#)

[PaperCitationContexts.txt](#)

[PaperFieldsOfStudy.txt](#)

[PaperLanguages.txt](#)

[PaperRecommendations.txt](#)

[PaperReferences.txt](#)

[PaperResources.txt](#)

[PaperUrls.txt](#)

[Papers.txt](#)

[RelatedFieldOfStudy.txt](#)

Documents the complete, most recent Microsoft Academic Graph entity data schema, including the name and type of each attribute.

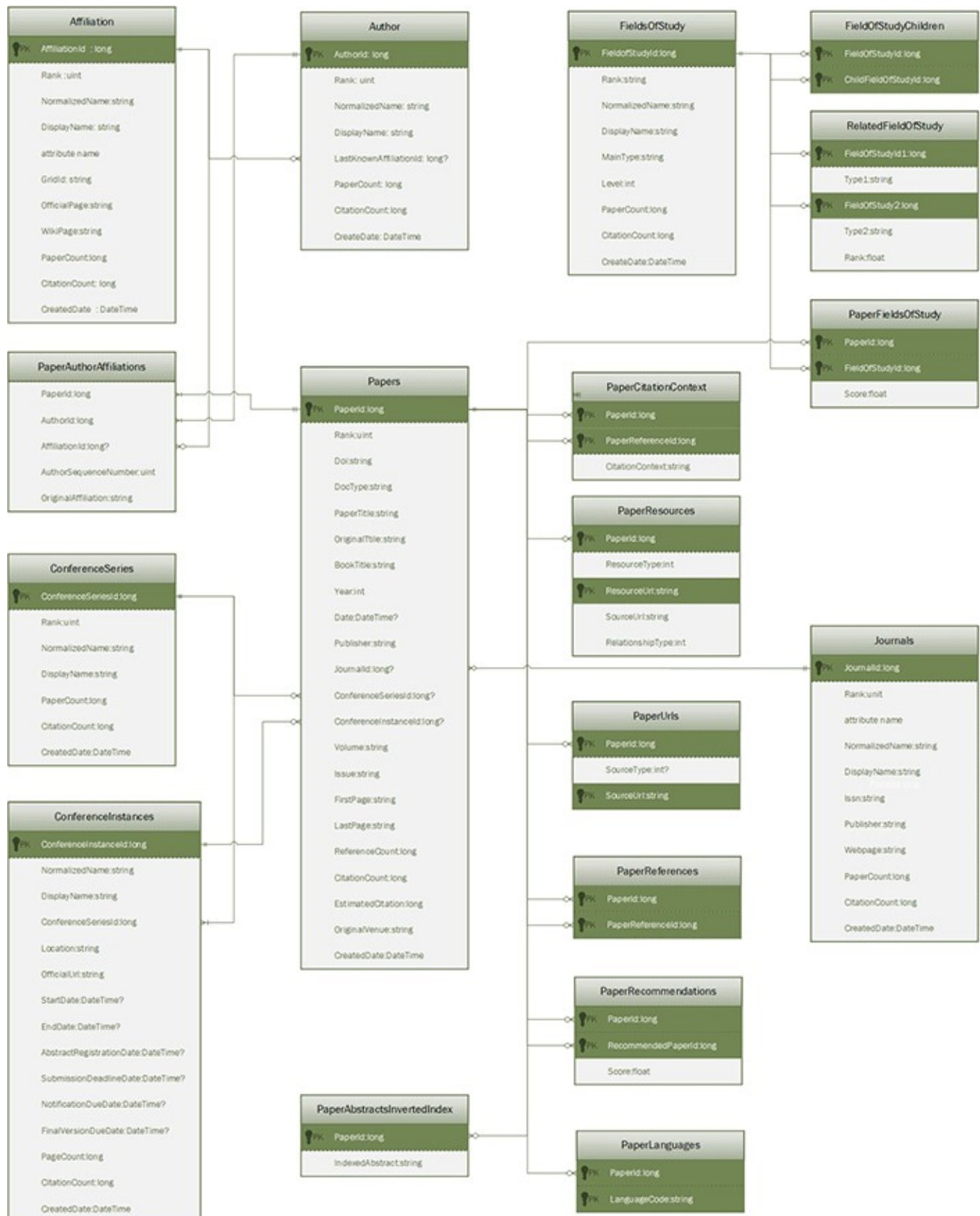
## Open Data License: **ODC-BY**

When using Microsoft Academic data (MAG, MAKES, etc.) in a product or service, or including data in a redistribution, please acknowledge Microsoft Academic using the URI <https://aka.ms/msracad>. For publications and reports, please cite the following article:

## ① Note

Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. 2015. An Overview of Microsoft Academic Service (MA) and Applications. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 243-246.  
DOI=<http://dx.doi.org/10.1145/2740908.2742839>

# Entity Relationship Diagram



## Note on "rank"

"Rank" values in the entity files are the log probability of an entity being important multiplied by a constant(-1000), i.e.:

### ⓘ Note

$\text{Rank} = -1000 * \ln(\text{probability of an entity being important})$

## Affiliations.txt

Column #	Name	Type	Note
1	AffiliationId	long	PRIMARY KEY
2	Rank	uint	
3	NormalizedName	string	
4	DisplayName	string	
5	GridId	string	
6	OfficialPage	string	
7	WikiPage	string	
8	PaperCount	long	
9	CitationCount	long	
10	CreatedDate	DateTime	

## Authors.txt

Column #	Name	Type	Note
----------	------	------	------

Column #	Name	Type	Note
1	AuthorId	long	PRIMARY KEY
2	Rank	uint	
3	NormalizedName	string	
4	DisplayName	string	
5	LastKnownAffiliationId	long?	
6	PaperCount	long	
7	CitationCount	long	
8	CreatedDate	DateTime	

## ConferenceInstances.txt

Column #	Name	Type	Note
1	ConferenceInstanceId	long	PRIMARY KEY
2	NormalizedName	string	
3	DisplayName	string	
4	ConferenceSeriesId	long	FOREIGN KEY REFERENCES ConferenceSeries (ConferenceSeriesId)
5	Location	string	
6	OfficialUrl	string	
7	StartDate	DateTime?	
8	EndDate	DateTime?	

Column #	Name	Type	Note
9	AbstractRegistrationDate	DateTime?	
10	SubmissionDeadlineDate	DateTime?	
11	NotificationDueDate	DateTime?	
12	FinalVersionDueDate	DateTime?	
13	PaperCount	long	
14	CitationCount	long	
15	CreatedDate	DateTime	

## ConferenceSeries.txt

Column #	Name	Type	Note
1	ConferenceSeriesId	long	PRIMARY KEY
2	Rank	uint	
3	NormalizedName	string	
4	DisplayName	string	
5	PaperCount	long	
6	CitationCount	long	
7	CreatedDate	DateTime	

## FieldsOfStudyChildren.txt

Column #	Name	Type	Note
----------	------	------	------

Column #	Name	Type	Note
1	FieldOfStudyId	long	FOREIGN KEY REFERENCES FieldsOfStudy(FieldOfStudyId)
2	ChildFieldOfStudyId	long	FOREIGN KEY REFERENCES FieldsOfStudy(FieldOfStudyId)

## FieldsOfStudy.txt

Column #	Name	Type	Note
1	FieldOfStudyId	long	PRIMARY KEY
2	Rank	uint	
3	NormalizedName	string	
4	DisplayName	string	
5	MainType	string	
6	Level	Int	0 - 5
7	PaperCount	long	
8	CitationCount	long	
9	CreatedDate	DateTime	

## Journals.txt

Column #	Name	Type	Note
1	JournalId	long	PRIMARY KEY
2	Rank	uint	
3	NormalizedName	string	
4	DisplayName	string	



Column #	Name	Type	Note
5	Issn	string	
6	Publisher	string	
7	Webpage	string	
8	PaperCount	long	
9	CitationCount	long	
10	CreatedDate	DateTime	

## PaperAbstractInvertedIndex.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	IndexedAbstract	string	See <a href="#">Microsoft Academic Graph FAQ</a>

## PaperAuthorAffiliations.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	AuthorId	long	FOREIGN KEY REFERENCES Authors(AuthorId)
3	AffiliationId	long?	FOREIGN KEY REFERENCES Affiliations(AffiliationId)
4	AuthorSequenceNumber	uint	1-based author sequence number. 1: the 1st author listed on paper.
5	OriginalAffiliation	string	



### ① Note

It is possible to have multiple rows with same (PaperId, AuthorId, AffiliationId) when an author is associated with multiple affiliations.

## PaperCitationContexts.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	PaperReferenceId	long	FOREIGN KEY REFERENCES Papers(PaperId)
3	CitationContext	string	

## PaperFieldsOfStudy.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	FieldOfStudyId	long	FOREIGN KEY REFERENCES FieldsOfStudy(FieldOfStudyId)
3	Score	float	Confidence range between 0 and 1. Bigger number representing higher confidence.

## PaperLanguages.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	LanguageCode	string	

## PaperRecommendations.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	RecommendedPaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
3	Score	float	Confidence range between 0 and 1. Bigger number representing higher confidence.

## PaperReferences.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	PaperReferenceId	long	FOREIGN KEY REFERENCES Papers(PaperId)

## PaperResources.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	ResourceType	int	
3	ResourceUrl	string	Bit flags. 1 = Project, 2 = Data, 4 = Code
4	SourceUrl	string	
5	RelationshipType	int	Bit flags. 1 = Own, 2 = Cite

## PaperUrls.txt

Column #	Name	Type	Note
1	PaperId	long	FOREIGN KEY REFERENCES Papers(PaperId)
2	SourceType	int?	
3	SourceUrl	string	

## Papers.txt

Column #	Name	Type	Note
1	PaperId	long	PRIMARY KEY
2	Rank	uint	
3	Doi	string	
4	DocType	string	Book, BookChapter, Conference, Journal, Patent, NULL : unknown
5	PaperTitle	string	
6	OriginalTitle	string	
7	BookTitle	string	
8	Year	int	
9	Date	DateTime?	
10	Publisher	string	
11	JournalId	long?	FOREIGN KEY REFERENCES Journals(JournalId)
12	ConferenceSeriesId	long?	FOREIGN KEY REFERENCES ConferenceSeries (ConferenceSeriesId)
13	ConferenceInstancelId	long?	

Column #	Name	Type	Note
			FOREIGN KEY REFERENCES ConferenceInstances (ConferenceInstanceld)
14	Volume	string	
15	Issue	string	
16	FirstPage	string	
17	LastPage	string	
18	ReferenceCount	long	
19	CitationCount	long	
20	EstimatedCitationCount	long	
21	OriginalVenue	string	
22	CreatedDate	DateTime	

## RelatedFieldOfStudy.txt

Column #	Name	Type	Note
1	FieldOfStudyId1	long	FOREIGN KEY REFERENCES FieldsOfStudy(FieldOfStudyId)
2	DisplayName 1	string	
3	Type1	string	disease, disease_cause, medical_treatment, symptom
4	FieldOfStudyId2	long	FOREIGN KEY REFERENCES FieldsOfStudy(FieldOfStudyId)
5	DisplayMame 2	string	
6	Type2	string	disease, disease_cause, medical_treatment, symptom

**Column**

#	Name	Type	Note
7	Rank	float	Confidence range between 0 and 1. Bigger number representing higher confidence.